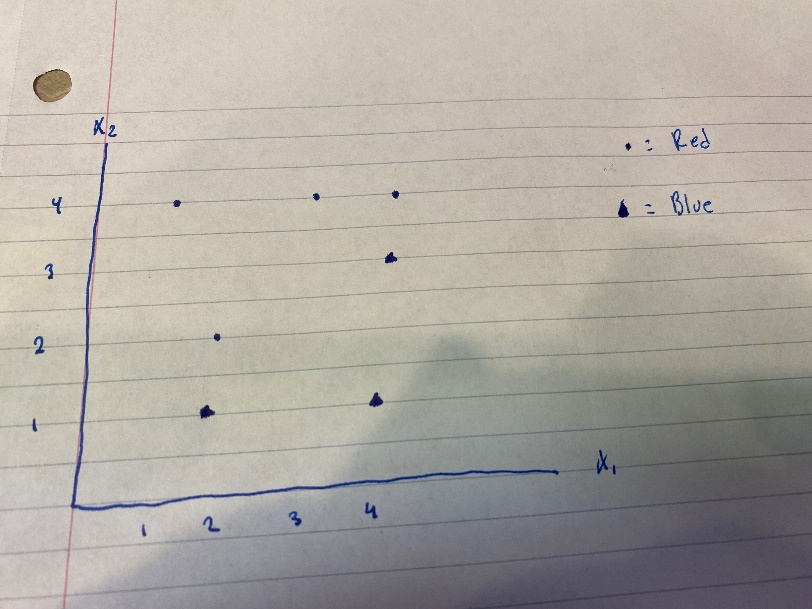
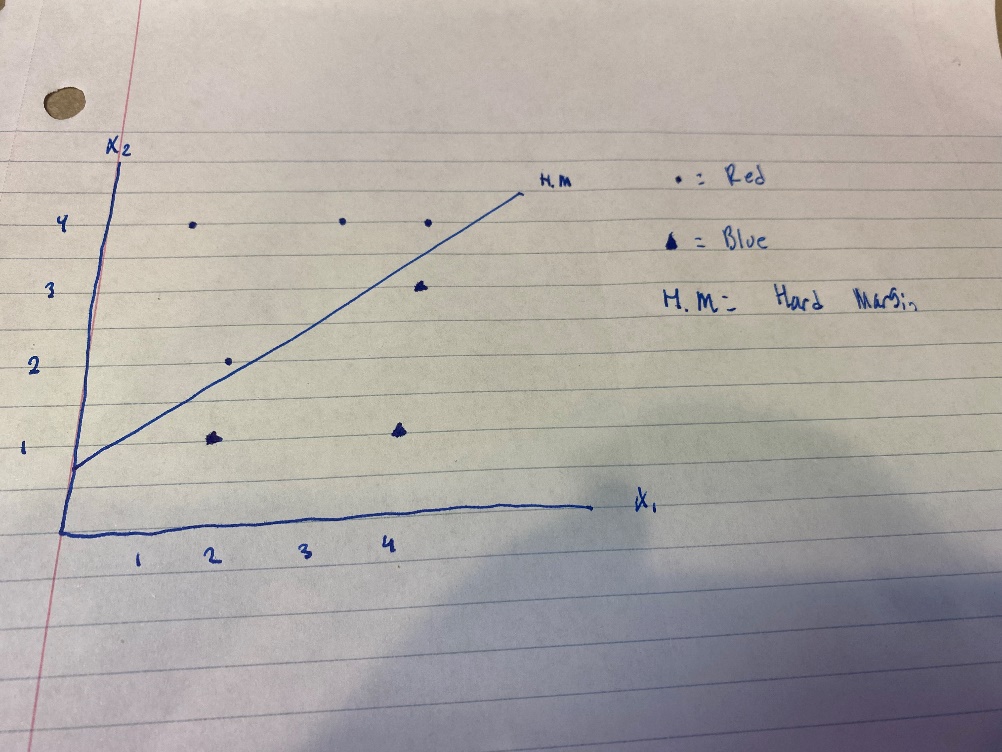
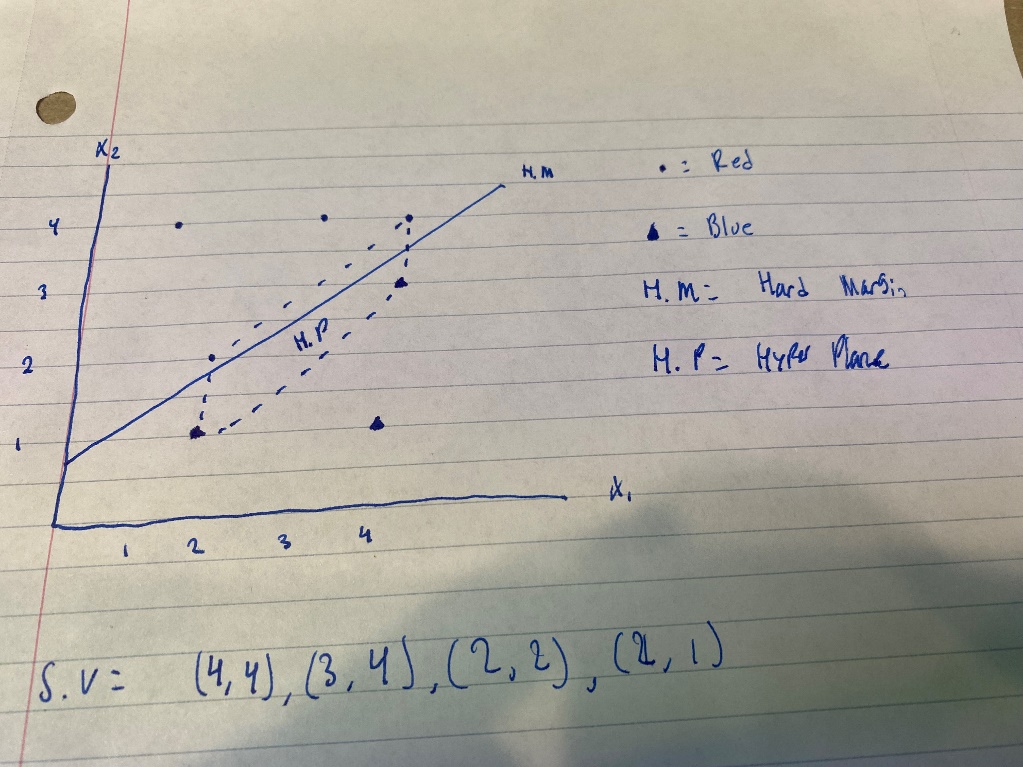
Homework 4 machine learning

1.

A: 

B: 

C: (4,4),(3,4),(2,2),(2,1)

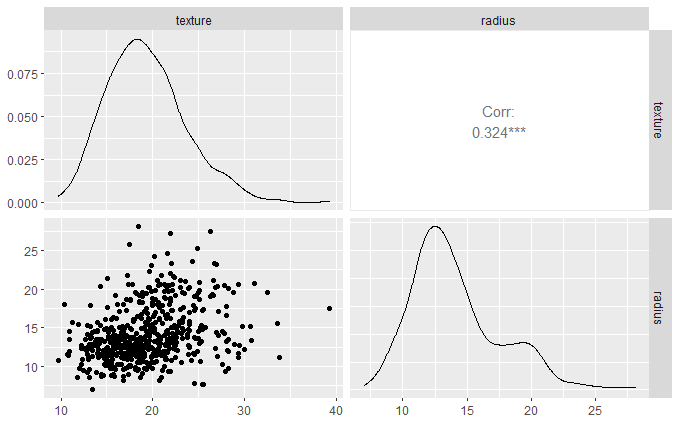
D: 

The equation of the hyper plane is: a(4,4)+b(3,4)+c(2,2)+d(2,1) + intercept of around .9

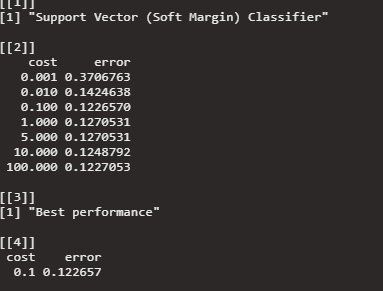
E: classify red if x2 is larger than x1 otherwise classify as blue.

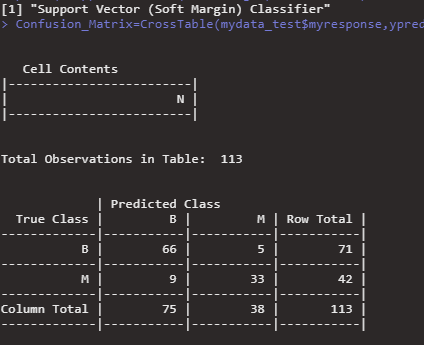
F: there should be no impact to the hyper plane, as the 7th value is not a support vector.

2.

A: there seems to be lots of overlap between texture and radius. It seems like there is a very slight postive correlation between texture and radius. If we were only to use these two variables, I think our predictions would not be very good. We would not be able to get a good hard classifier since there is significant overlap between them. Soft margin classifier would also fail for this as there would be many violations that occur.

B: the best cost parameter is .1

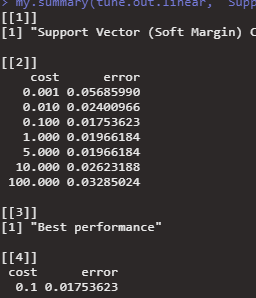


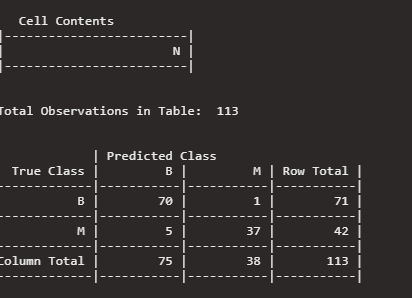


The model over estimated the number of bengin tumors while under estimating the number of malignant tumors. The error rate was decent at 12%. The accuracy rate for benign is .9295 while the accuracy rate for malignant is .7857. the mode is signficantly better at predicting benign tumors however we can evenly weigh these outcomes as missing an malignant tumor is extremely more dangerous than predicting a benign tumor as malignant.

C:

The cost that allows us the lowest error is .1





The overall accuracy of this model is significantly better than the first with an error rate of about 1.7%. This model is very good at predicting benign tumors with an accuracy of .9859 which is significantly better than our previous model. This models accuracy for malignant tumors was also significantly better than our previous model as well. This model had an accuracy of .8809. This model is miles better than the other however it is unbalanced in terms of specificity and sensitivity.